

# FAITHFULNESS OF PROBABILITY DISTRIBUTIONS AND GRAPHS

BY KAYVAN SADEGHI

*University of Cambridge*

A main question in graphical models and causal inference is whether, given a probability distribution  $P$  (which is usually an underlying distribution of data), there is a graph (or graphs) to which  $P$  is faithful. The main goal of this paper is to provide a theoretical answer to this problem. We work with general independence models, which contain probabilistic independence models as a special case. We exploit a generalization of ordering, called preordering, of the nodes of (mixed) graphs. This allows us to provide sufficient conditions for a given independence model to be Markov to a graph with the minimum possible number of edges, and more importantly, necessary and sufficient conditions for a given probability distribution to be faithful to a graph. We present our results for the general case of mixed graphs, but specialize the definitions and results to the better-known subclasses of undirected (concentration) and bidirected (covariance) graphs as well as directed acyclic graphs.

**1. Introduction.** Graphs have been used in graphical models in order to capture the conditional independence structure of probability distributions. Generally speaking, nodes of the graph correspond to random variables and edges to conditional dependencies [17]. The connection between graphs and probability distributions is usually established in the literature by the concept of Markov property [4; 27; 40], which ensures that if there is a specific type of separation between nodes  $i$  and  $j$  of the graph “given the node subset  $C$ ” then random variables  $X_i$  and  $X_j$  are conditionally independent given the random vector  $X_C$  in the probability distribution. However, the “ultimate” connection between probability distributions and graphs requires the other direction to hold, namely for every conditional independence in the probability distribution to correspond to a separation in the graph. This connection has been called faithfulness of the probability

---

\*Partly supported by grant #FA9550-12-1-0392 from the U.S. Air Force Office of Scientific Research (AFOSR) and the Defense Advanced Research Projects Agency (DARPA).

AMS 2000 subject classifications: Primary 62H99; secondary 62A99

Keywords and phrases: causal discovery, compositional graphoid, directed acyclic graph, faithfulness, graphical model selection, independence model, Markov property, mixed graph, structural learning

distribution and the graph [38], and the graph has been called the perfect map of such a distribution [27].

However, “given a probability distribution  $P$  whether there is a graph (or graphs) to which  $P$  is faithful” is an open problem, and consequently so is the problem of finding these graphs. This problem can be raised for any type of graph existing in the literature of graphical models ranging from different types of mixed graphs with three types of edges [32; 45; 34] and different types of chain graphs [20; 10; 2; 5; 45] to better-known classes of undirected (concentration) [6] and bidirected (covariance) [15] graphs as well as directed acyclic graphs [16; 27]. Our goal is to provide an answer to this problem. A similar problem of “given a graph whether there is a family of distributions faithful to it” has been answered for very specific types of graphs and specific types of distributions; e.g., for Gaussian distributions and undirected graphs [21], DAGs [27; 11; 38], ancestral graphs [32], and LWF chain graphs [25]; and discrete distributions and DAGs [11; 22] and LWF chain graphs [24].

The concept of faithfulness was originally defined for the purpose of causal inference [38; 27], and the theory developed in this paper can be interpreted in causal language. A main approach to causal inference is based on graphical representations of causal structures, usually represented by *causal graphs* that are directed acyclic with nodes as random variables (i.e., a Bayesian network) that are assumed to capture the true causal structure; see, e.g., [23]. A main assumption made is called the *causal faithfulness condition* (CFC) stating that the true probability distribution is faithful to the true causal structure, similar to what was discussed above in a more general setting. For an extensive discussion on the CFC, see [50], and for related philosophical discussions, see, e.g., [49; 39]. Although the results in this paper cover those of causal Bayesian networks, we have developed our theory to a much more general case of simultaneous representation of “direct effects”, “confounding”, and “non-causal symmetric dependence structures”; see, e.g., [28].

In this paper, we work with the general independence model  $\mathcal{J}$ . Independence models  $\mathcal{J}(P)$  induced by probability distributions  $P$  are a special case. We provide necessary and sufficient conditions for  $\mathcal{J}$  to be graphical, i.e. to be faithful to a graph. In short, as proved in Theorem 2,  $\mathcal{J}$  is graphical if and only if it satisfies the so-called compositional graphoid axioms as well as singleton-transitivity, and what we call ordered upward- and downward-stability.

As apparent from their names, ordered upward- and downward-stability depend on a generalization of ordering of variables, and consequently the nodes of the graph (called preordering). We provide our results for the

most general case of mixed graphs, which contains almost all classes of graphs used in graphical models as subclasses (with the exception of the chain graphs with the AMP Markov property [2] and its generalizations [26]). However, based on the preordering w.r.t. which ordered upward- and downward-stability are satisfied, one can deduce to what type (or types) of graph  $P$  is faithful.

Since many of the subclasses of mixed graphs are not as well-known or used as some simpler classes of graphs, we provide a specialization of definitions and the main results for undirected and bidirected graphs as well as directed acyclic graphs and for solely probability distributions at the end of the paper. We advise the readers that are only familiar with or interested in these simpler subclasses to skip the general definitions and results for mixed graphs, and focus on the specialization.

The structure of the paper is as follows: In the next section, we provide the basic definitions for graphs as well as different classes of graphs in graphical models, and independence models. In Section 3, we define and provide basic results for Markovness and faithfulness of independence models and graphs. In Section 4, we define and exploit the preordering for (so-called antierial) graphs, and then we define the concept of upward- and downward-stability. In Section 5, we provide sufficient conditions for an independence model to be so-called minimally Markov to a graph, and then we provide necessary and sufficient conditions for an independence model to be faithful to a graph. In Section 6, we specialize the definitions and results to the classes of undirected and bidirected graphs as well as directed acyclic graphs and for probability distributions. In Section 7, we show how other results in the literature concerning faithfulness of certain probability distributions and certain classes of graphs are corollaries of our results. This also provides a nice set of examples for the theory presented in the paper. In Section 8, we end the paper with a short summary and outline future work.

## 2. Definitions and concepts.

*2.1. Graph theoretical definitions and concepts.* A *graph*  $G$  is a triple consisting of a *node* set or *vertex* set  $V$ , an *edge* set  $E$ , and a relation that with each edge associates two nodes (not necessarily distinct), called its *endpoints*. When nodes  $i$  and  $j$  are the endpoints of an edge, they are *adjacent*. We call a node adjacent to the node  $i$ , a *neighbour* of  $i$ , and denote the set of all neighbours of  $i$  by  $\text{ne}(i)$ . We say that an edge is *between* its two endpoints. We usually refer to a graph as an ordered pair  $G = (V, E)$ . Graphs  $G_1 = (V_1, E_1)$  and  $G_2 = (V_2, E_2)$  are called *equal* if  $(V_1, E_1) = (V_2, E_2)$ . In this case we write  $G_1 = G_2$ .

Notice that our graphs are *labeled*, i.e. every node is considered a different object. Hence, for example, graph  $i \text{ --- } j \text{ --- } k$  is not equal to  $j \text{ --- } i \text{ --- } k$ .

A *loop* is an edge with the same endpoints. Here we only discuss graphs without loops. *Multiple edges* are edges with the same pair of endpoints. A *simple graph* has neither loops nor multiple edges.

Graphs in this paper are so-called *mixed graphs*, which contain three types of edges: *lines* ( $i \text{ --- } j$ ), *arrows* ( $i \rightarrow j$ ), and *arcs* ( $i \leftrightarrow j$ ), where in the two latter cases, we say that there is an *arrowhead* at  $j$ . If we remove all arrowheads from edges of the graph  $G$ , the obtained undirected graph is called the *skeleton* of  $G$ . The simple graph whose edge  $ij$  indicates whether there is an edge (or multiple edges) between  $i$  and  $j$  in  $G$  is called the *adjacency graph* of  $G$ . It is clear that if a graph is simple then its adjacency graph is the same as its skeleton. Here our results only concern simple graphs; hence, we denote the simple skeleton by  $\text{sk}(G)$ . We say that we *direct* the edges of a skeleton by putting arrowheads at the edges in order to obtain mixed graphs.

A *subgraph* of a graph  $G_1$  is graph  $G_2$  such that  $V(G_2) \subseteq V(G_1)$  and  $E(G_2) \subseteq E(G_1)$  and the assignment of endpoints to edges in  $G_2$  is the same as in  $G_1$ . We define a subgraph *induced by edges*  $A \subseteq E$  of  $G = (V, E)$  to be a subgraph that contains  $V$  as the node set and all and only edges in  $A$  as the edge set.

A *walk* is a list  $\langle v_0, e_1, v_1, \dots, e_k, v_k \rangle$  of nodes and edges such that for  $1 \leq i \leq k$ , the edge  $e_i$  has endpoints  $v_{i-1}$  and  $v_i$ . A *path* is a walk with no repeated node or edge. A maximal set of nodes in a graph whose members are connected by some paths constitutes a *connected component* of the graph. A *cycle* is a walk with no repeated nodes or edges except for  $v_0 = v_k$ .

We say a walk is *between* the first and the last nodes of the list in  $G$ . We call the first and the last nodes *endpoints* of the walk and all other nodes *inner nodes*.

A *subwalk* or a *subpath* of a walk  $\omega = \langle e_1, \dots, e_n \rangle$  is any subsequence  $\langle e_{n_1}, e_{n_2}, \dots, e_{n_p} \rangle$  of  $\omega$  which forms a walk or path. However, we normally express a subwalk by a list of nodes when it is apparent from the context.

A walk  $\omega = \langle i = i_0, i_1, \dots, i_n = j \rangle$  is *directed* from  $i$  to  $j$  if all edges  $i_k i_{k+1}$ ,  $0 \leq k \leq n-1$ , are arrows pointing from  $i_k$  to  $i_{k+1}$ . If there is a directed walk from  $i$  to  $j$  then  $i$  is an *ancestor* of  $j$  and  $j$  is a *descendant* of  $i$ . We denote the set of ancestors of  $j$  by  $\text{an}(j)$ .

A walk  $\omega = \langle i = i_0, i_1, \dots, i_n = j \rangle$  is *semi-directed* from  $i$  to  $j$  if it has at least one arrow, no arcs, and every arrow  $i_k i_{k+1}$  is pointing from  $i_k$  to  $i_{k+1}$ . A walk between  $i$  and  $j$  is *anterior* from  $i$  to  $j$  if it is semi-directed from  $i$  to  $j$  or if it is undirected. If there is an anterior walk from  $i$  to  $j$  then we

also say that  $i$  is an *anterior* of  $j$ . We use the notation  $\text{ant}(j)$  for the set of all antiors of  $j$ . For a set  $A$ , we define  $\text{ant}(A) = \bigcup_{j \in A} \text{ant}(j) \setminus A$ .

Notice that, unlike in most places in the literature (e.g. [32]), we use walks instead of paths to define ancestors and antiors. Because of this and the fact that ancestral graphs have no arrowheads pointing to lines, our definition of anterior extends the notion of anterior for ancestral graphs in [32] with the modification that in this paper, a node is not an anterior of itself. Using walks instead of paths is immaterial, as shown in [18].

A *section*  $\rho$  of a walk is a maximal subwalk consisting only of lines, meaning that there is no other subwalk that only consists of lines and includes  $\rho$ . Thus, any walk decomposes uniquely into sections; these are not necessarily edge-disjoint and sections may also be single nodes. A section  $\rho$  on a walk  $\omega$  is called a *collider section* if one of the following walks is a subwalk of  $\omega$ :  $i \rightarrow \rho \leftarrow j$ ,  $i \leftrightarrow \rho \leftarrow j$ ,  $i \leftrightarrow \rho \leftrightarrow j$ . All other sections on  $\omega$  are called *non-collider* sections. Notice that a section may be a collider on one walk and a non-collider on another, but we may speak of collider or non-collider sections without mentioning the relevant walk when this is apparent from the context.

**2.2. Different classes of graphs.** All the subclasses of graphs included here are *acyclic* in the sense that they do not contain semi-directed cycles. The most general class of graphs discussed here is the class of *chain mixed graphs* (CMGs) [34] that contains all mixed graphs without semi-directed cycles. They may have multiple edges containing an arc and a line or an arc and an arrow but not a combination of an arrow and a line or arrows in opposite directions, as such combinations would constitute semi-directed cycles. In this paper, in using the term “graph” we mean a CMG unless otherwise stated.

A general class of graphs that plays an important role in this paper is the class of *antierial graphs* (AnGs) [34]. AnGs are CMGs in which an endpoint of an arc cannot be an anterior of the other endpoint.

CMGs include *summary graphs* (SGs) [45] as a subclass, and AnGs include *ancestral graphs* (AGs) [32] and *acyclic directed mixed graphs* (ADMGs) [31]. These have typically been introduced to describe independence structures obtained by marginalization and conditioning in DAG independence models; see for example [33].

AnGs also contain *undirected and bidirected chain graphs* (CGs). A chain graph is an acyclic graph so that if we remove all arrows, all connected components of the resulting graph – called *chain components* – contain one type of edge only. If all chain components contain lines, the chain graph is an

undirected chain graph (UCG) (known as *LWF chain graphs*); if all chain components contain arcs it is a bidirected chain graph (BCG) (known as *multivariate regression chain graphs*). AnGs also contain *Regression graphs* [48], which are chain graphs consisting of lines and arcs (although dashed undirected edges have mostly been used instead of arcs in the literature), where there is no arrowhead pointing to lines.

These also contain graphs with only one type of edge; namely *undirected graphs* (UGs), containing only lines; *bidirected graphs* (BGs), containing only bidirected edges; and *directed acyclic graphs* (DAGs), containing only arrows and being acyclic. Clearly, a graph without arrows has no semi-directed cycles, and a semi-directed cycle in a graph with only arrows is a directed cycle. [5; 15; 46; 8] used the terms *concentration graphs* and *covariance graphs* for UGs and BGs, referring to their independence interpretation associated with covariance and concentration matrices for Gaussian graphical models. DAGs have been particularly useful to describe causal Markov relations; see for example [16; 27; 19; 11; 38].

For an extensive discussion on the subclasses of acyclic graphs and their relationships and hierarchy, see [18].

**2.3. Independence models and their properties.** An *independence model*  $\mathcal{J}$  over a finite set  $V$  is a set of triples  $\langle X, Y \mid Z \rangle$  (called *independence statements*), where  $X$ ,  $Y$ , and  $Z$  are disjoint subsets of  $V$ ;  $Z$  may be empty, but  $\langle \emptyset, Y \mid Z \rangle$  and  $\langle X, \emptyset \mid Z \rangle$  are always included in  $\mathcal{J}$ . The independence statement  $\langle X, Y \mid Z \rangle$  is read as “ $X$  is independent of  $Y$  given  $Z$ ”. Independence models may in general have a probabilistic interpretation, but not necessarily. Similarly, not all independence models can be easily represented by graphs. For further discussion on general independence models, see [41].

In order to define probabilistic independence models, consider a set  $V$  and a collection of random variables  $\{X_\alpha\}_{\alpha \in V}$  with state spaces  $\mathcal{X}_\alpha$ ,  $\alpha \in V$  and joint distribution  $P$ . We let  $X_A = \{X_v\}_{v \in A}$  etc. for each subset  $A$  of  $V$ . For disjoint subsets  $A$ ,  $B$ , and  $C$  of  $V$  we use the short notation  $A \perp\!\!\!\perp B \mid C$  to denote that  $X_A$  is *conditionally independent of  $X_B$  given  $X_C$*  [7; 17], i.e. that for any measurable  $\Omega \subseteq \mathcal{X}_A$  and  $P$ -almost all  $x_B$  and  $x_C$ ,

$$P(X_A \in \Omega \mid X_B = x_B, X_C = x_C) = P(X_A \in \Omega \mid X_C = x_C).$$

We can now induce an independence model  $\mathcal{J}(P)$  by letting

$$\langle A, B \mid C \rangle \in \mathcal{J}(P) \text{ if and only if } A \perp\!\!\!\perp B \mid C \text{ w.r.t. } P.$$

Similarly we use the notation  $A \not\perp\!\!\!\perp B \mid C$  for  $\langle A, B \mid C \rangle \notin \mathcal{J}(P)$ .

In order to use graphs to represent independence models, the notion of *separation* in a graph is fundamental. For three disjoint subsets  $A$ ,  $B$ , and  $C$ , we use the notation  $A \perp B | C$  if  $A$  and  $B$  are *separated* given  $C$ , and  $A \not\perp B | C$  for  $A$  and  $B$  not separated given  $C$ .

The separation can be unified under one definition for all graphs by using walks instead of paths: A walk  $\pi$  is *connecting* given  $C$  if every collider section of  $\pi$  has a node in  $C$  and all non-collider sections are disjoint from  $C$ . For pairwise disjoint subsets  $(A, B, C)$ ,  $A \perp B | C$  if there are no connecting walks between  $A$  and  $B$  given  $C$ . This is in wording the same as the definition in [42] for undirected (LWF) chain graphs (although it is in fact a generalization as collider sections are generalized), and a generalization of the *m-separation* used with different wordings in [32; 45; 48] for AGs and SGs, and of the *d-separation* of [27]. For UGs, the notion has a direct intuitive meaning, so that  $A \perp B | C$  if all paths from  $A$  to  $B$  pass through  $C$ .

If  $A$ ,  $B$ , or  $C$  has only one member  $\{i\}$ ,  $\{j\}$ , or  $\{k\}$ , for better readability, we write  $\langle i, j | k \rangle \in \mathcal{J}$  instead of  $\langle \{i\}, \{j\} | \{k\} \rangle \in \mathcal{J}$ ; and similarly for  $i \perp j | k$  and  $i \perp\!\!\!\perp j | k$ . We also write  $A \perp B$  when  $C = \emptyset$ ; and similarly  $A \perp\!\!\!\perp B$ .

A graph  $G$  induces an independence model  $\mathcal{J}(G)$  by separation, letting

$$\langle A, B | C \rangle \in \mathcal{J}(G) \iff A \perp B | C \text{ in } G.$$

An independence model  $\mathcal{J}$  over a set  $V$  is a *semi-graphoid* if it satisfies the four following properties for disjoint subsets  $A$ ,  $B$ ,  $C$ , and  $D$  of  $V$ :

1.  $\langle A, B | C \rangle \in \mathcal{J}$  if and only if  $\langle B, A | C \rangle \in \mathcal{J}$  (*symmetry*);
2. if  $\langle A, B \cup D | C \rangle \in \mathcal{J}$  then  $\langle A, B | C \rangle \in \mathcal{J}$  and  $\langle A, D | C \rangle \in \mathcal{J}$  (*decomposition*);
3. if  $\langle A, B \cup D | C \rangle \in \mathcal{J}$  then  $\langle A, B | C \cup D \rangle \in \mathcal{J}$  and  $\langle A, D | C \cup B \rangle \in \mathcal{J}$  (*weak union*);
4. if  $\langle A, B | C \cup D \rangle \in \mathcal{J}$  and  $\langle A, D | C \rangle \in \mathcal{J}$  then  $\langle A, B \cup D | C \rangle \in \mathcal{J}$  (*contraction*).

Notice that the reverse implication of contraction clearly holds by decomposition and weak union. A semi-graphoid for which the reverse implication of the weak union property holds is said to be a *graphoid*; that is, it also satisfies

5. if  $\langle A, B | C \cup D \rangle \in \mathcal{J}$  and  $\langle A, D | C \cup B \rangle \in \mathcal{J}$  then  $\langle A, B \cup D | C \rangle \in \mathcal{J}$  (*intersection*).

Furthermore, a graphoid or semi-graphoid for which the reverse implication of the decomposition property holds is said to be *compositional*, that is, it also satisfies



6. if  $\langle A, B \mid C \rangle \in \mathcal{J}$  and  $\langle A, D \mid C \rangle \in \mathcal{J}$  then  $\langle A, B \cup D \mid C \rangle \in \mathcal{J}$  (*composition*).

Separation in graphs satisfy all of these properties; see Theorem 1 in [18]:

PROPOSITION 1. *For any graph  $G$ , the independence model  $\mathcal{J}(G)$  is a compositional graphoid.*

On the other hand, probabilistic independence models are always semi-graphoids [27], whereas the converse is not necessarily true; see [40]. If, for example,  $P$  has strictly positive density, the induced independence model is always a graphoid; see e.g. Proposition 3.1 in [17]. See also [29] for a necessary and sufficient condition for the intersection property to hold. If the distribution  $P$  is a regular multivariate Gaussian distribution,  $\mathcal{J}(P)$  is a compositional graphoid; e.g. see [41]. Probabilistic independence models with positive densities are not in general compositional; this only holds for special types of multivariate distributions such as, for example, Gaussian distributions and the symmetric binary distributions used in [47].

Another important property that is satisfied by separation in all graphs, but not necessarily for probabilistic independence models, is *singleton-transitivity* (also called *weak transitivity* in [27], where it is shown that for Gaussian and binary distributions  $P$ ,  $\mathcal{J}(P)$  always satisfies it). For  $i, j$ , and  $k$ , single elements in  $V$ ,

7. if  $\langle i, j \mid C \rangle \in \mathcal{J}$  and  $\langle i, j \mid C \cup \{k\} \rangle \in \mathcal{J}$  then  $\langle i, k \mid C \rangle \in \mathcal{J}$  or  $\langle j, k \mid C \rangle \in \mathcal{J}$  (singleton-transitivity).

A singleton-transitive compositional graphoid is equivalent to what is called a *Gaussoid* in [21] with a rather different axiomatization. The name reminds one that these are the axioms satisfied by the independence model of a regular Gaussian distribution [27].

PROPOSITION 2. *For a graph  $G$ ,  $\mathcal{J}(G)$  satisfies singleton-transitivity.*

PROOF. If there is a walk between  $i$  and  $k$  given  $C$  and a walk between  $j$  and  $k$  given  $C$  then by connecting these two walks, we obtain a walk  $\pi$  between  $i$  and  $j$ . Except the section  $\rho$  that contains  $k$ , all collider sections on  $\pi$  have a node in  $C$  and all non-collider sections are outside  $C$ . Hence depending on  $\rho$  being a collider or non-collider,  $\pi$  is connecting given  $C \cup \{k\}$  or  $C$  respectively.  $\square$

### 3. Markov and faithful independence models.



3.1. *Markov properties.* For a graph  $G = (V, E)$ , an independence model  $\mathcal{J}$  defined over  $V$  satisfies the *global Markov property* w.r.t.  $G$ , or is simply *Markov* to  $G$ , if for disjoint subsets  $A$ ,  $B$ , and  $C$  of  $V$  it holds that

$$A \perp B | C \implies \langle A, B | C \rangle \in \mathcal{J},$$

or equivalently  $\mathcal{J}(G) \subseteq \mathcal{J}$ . In particular, the graphical independence  $\mathcal{J}(G)$  is trivially Markov to its graph  $G$ .

For a probability distribution  $P$ , we simply say  $P$  is *Markov* to a graph  $G$  if  $\mathcal{J}(P)$  is Markov to  $G$ , i.e., if  $\mathcal{J}(G) \subseteq \mathcal{J}(P)$ . Notice that every independence model over  $V$  is Markov to the complete graph with the node set  $V$ .

For a graph  $G = (V, E)$ , an independence model  $\mathcal{J}$  defined over  $V$  satisfies a *pairwise Markov property* w.r.t. a graph  $G$ , or is simply *pairwise Markov* to  $G$ , if for every non-adjacent pair of nodes  $i, j$ , it holds that  $\langle i, j | C(i, j) \rangle \in \mathcal{J}$ , for some  $C(i, j)$ , where  $C(i, j)$  is the *conditioning set* of the pairwise Markov property and does not include  $i$  and  $j$ .

The independence model  $\mathcal{J}(G)$  only satisfies pairwise Markov properties w.r.t. what is called *maximal graphs*, which are graphs where the lack of an edge between  $i$  and  $j$  corresponds to a conditional separation statement for  $i$  and  $j$ . For a probability distribution  $P$ , we say  $P$  is *pairwise Markov* to a graph  $G$  if  $\mathcal{J}(P)$  is pairwise Markov to  $G$ .

For maximal graphs, a pairwise Markov property is defined by letting  $C(i, j) = \text{ant}(i) \cup \text{ant}(j) \setminus \{i, j\}$  [18], which we henceforth use as “the” pairwise Markov property. The conditioning set of the pairwise Markov property simplifies for DAGs,  $C(i, j) = \text{an}(i) \cup \text{an}(j) \setminus \{i, j\}$  [32; 35]; for BGs,  $C(i, j) = \emptyset$  (as defined in [46]); and for connected UGs,  $C(i, j) = V \setminus \{i, j\}$  (as defined in [17]).

This pairwise Markov property, is in fact equivalent to the global Markov property under compositional graphoids; see Theorem 4 of [18]:

**PROPOSITION 3.** *Let  $G$  be a maximal graph. If the independence model  $\mathcal{J}$  is a compositional graphoid, then  $\mathcal{J}$  is pairwise Markov to  $G$  if and only if  $\mathcal{J}$  is Markov to  $G$ .*

In particular, for UGs, the equivalence of the global and pairwise Markov properties holds under graphoids, and for BGs under compositional semi-graphoids.

Two undirected or two bidirected graphs  $G$  and  $H$ , where  $G \neq H$ , induce different independence models; i.e.  $\mathcal{J}(G) \neq \mathcal{J}(H)$ . This is not necessarily true for larger subclasses. We call two graphs  $G$  and  $H$  such that  $\mathcal{J}(G) = \mathcal{J}(H)$  *Markov equivalent*. Conditions for Markov equivalence for

most subclasses of graphs are known; see [44; 1; 48]. Notice that two Markov equivalent maximal graphs have the same skeleton.

**3.2. Faithfulness and minimal Markovness.** We say that an independence model  $\mathcal{J}$  and a probability distribution  $P$  are *faithful* if  $\mathcal{J} = \mathcal{J}(P)$ . Similarly, we say that  $\mathcal{J}$  and a graph  $G$  are faithful if  $\mathcal{J} = \mathcal{J}(G)$ . We also say that  $P$  and  $G$  are faithful if  $\mathcal{J}(P) = \mathcal{J}(G)$ . If  $P$  and  $G$  are faithful then we may sometimes also say that  $P$  is *faithful to*  $G$  or vice versa, although in principle faithfulness is a symmetric relation; the same holds for saying that  $\mathcal{J}$  is faithful to  $G$  or  $P$ . Notice that we are extending the definition of faithfulness to include the relation between independence models and graphs as well as independence models and probability distributions rather than only that between graphs and probability distributions as originally used in [38].

Thus, if  $\mathcal{J}$  and  $G$  are faithful then  $\mathcal{J}$  is Markov to  $G$ , but it also requires every independence statement to correspond to a separation in  $G$ . We say that  $\mathcal{J}$  or  $G$  is *probabilistic* if there is a distribution  $P$  that is faithful to  $\mathcal{J}$  or  $G$  respectively. If there is a graph that is faithful to  $\mathcal{J}$  or  $P$  then we say that  $\mathcal{J}$  or  $P$  is *graphical*.

Our main goal is to characterize graphical probability distributions, and in addition, if existent, to provide graphs that are faithful to a given  $P$ . We solve this problem for the general case of independence models  $\mathcal{J}$ .

For a given independence model  $\mathcal{J}$ , we define the *skeleton* of  $\mathcal{J}$ , denoted by  $\text{sk}(\mathcal{J})$ , to be the undirected graph that is obtained from  $\mathcal{J}$  as follows: we define the node set of  $\text{sk}(\mathcal{J})$  to be  $V$ , and for every pair of nodes  $i, j$ , we check whether  $\langle i, j | C \rangle \in \mathcal{J}$  holds for some  $C \subseteq V \setminus \{i, j\}$ ; if it does not then we draw an edge between  $i$  and  $j$ . One can similarly define  $\text{sk}(P)$  for a probability distribution  $P$  by checking whether  $i \perp\!\!\!\perp j | C$ . This is the same as the graph obtained by the first step of the SGS algorithm [12].

For UGs and BGs, the skeleton of  $\mathcal{J}$  uniquely determines the graph, whereas for other subclasses, there are several graphs with the same skeleton  $\text{sk}(\mathcal{J})$ . As will be seen, the preordering, defined in the next section, enables us to direct the edges of the skeleton.

**PROPOSITION 4.** *If an independence model  $\mathcal{J}$  is Markov to a graph  $G$  then  $\text{sk}(\mathcal{J})$  is a subgraph of  $\text{sk}(G)$ .*

**PROOF.** Suppose that there is no edge between  $i$  and  $j$  in  $\text{sk}(G)$ . Since the global Markov property w.r.t.  $G$  implies the pairwise Markov property, it holds that  $\langle i, j | C(i, j) \rangle \in \mathcal{J}$ , which implies  $i$  is not adjacent to  $j$  in  $\text{sk}(\mathcal{J})$ .  $\square$

Hence, if  $\mathcal{J}$  is Markov to a graph  $G$  such that  $\text{sk}(G) = \text{sk}(\mathcal{J})$  then  $G$  has the fewest number of edges among those to which  $\mathcal{J}$  is Markov. We say that  $\mathcal{J}$  is *minimally Markov* to a graph  $G$  if  $\mathcal{J}$  is Markov to  $G$  and  $\text{sk}(G) = \text{sk}(\mathcal{J})$ . The same can be defined for probability distributions, and has been used in the literature under the name of *(causal) minimality assumption* [28; 38; 50; 23]. Minimally Markov independence models to a graph are important since only these can also be faithful to the graph:

**PROPOSITION 5.** *If an independence model  $\mathcal{J}$  and a graph  $G$  are faithful then  $\mathcal{J}$  is minimally Markov to  $G$ .*

**PROOF.** Since  $\mathcal{J}$  is Markov to  $G$ , we need to prove that  $\text{sk}(G) = \text{sk}(\mathcal{J})$ . By Proposition 4,  $\text{sk}(\mathcal{J})$  is a subgraph of  $\text{sk}(G)$ . Now, suppose that there is no edge between  $i$  and  $j$  in  $\text{sk}(\mathcal{J})$ . By the construction of  $\text{sk}(\mathcal{J})$ , it holds that  $\langle i, j \mid C \rangle \in \mathcal{J}$  for some  $C$ . Since  $\mathcal{J}$  and  $G$  are faithful,  $i \perp j \mid C$  in  $G$ . This implies that  $i$  is not adjacent to  $j$  in  $G$ .  $\square$

Hence, we need to discuss conditions under which  $\mathcal{J}$  is minimally Markov to a graph as well as conditions for faithfulness of minimally Markov independence models and graphs. These will be presented in Section 5 after we define the preordering for sets and its validity for nodes of the graphs.

#### 4. Preordering in graphs and ordered stabilities.

4.1. *Preordering of the nodes of a graph.* Over a set  $V$ , a *partial order* is a binary relation  $\leq$  that satisfies the following properties for all members  $a, b, c \in V$ :

- $a \leq a$  (*reflexivity*);
- if  $a \leq b$  and  $b \leq a$  then  $a = b$  (*antisymmetry*);
- if  $a \leq b$  and  $b \leq c$  then  $a \leq c$  (*transitivity*).

If  $a \leq b$  or  $b \leq a$  then  $a$  and  $b$  are *comparable*; otherwise they are *incomparable*. A partial order under which every pair of elements is comparable is called a *total order* or a *linear order*.

A *preorder*  $\lesssim$  over  $V$ , on the other hand, is a binary relation that is only reflexive and transitive. Given a preorder  $\lesssim$  over  $V$ , one may define an equivalence relation  $\sim$  on  $V$  such that  $a \sim b$  if and only if  $a \lesssim b$  and  $b \lesssim a$ . This explains the use of the notation  $\lesssim$ . It then holds that

- if  $a \lesssim b$  and  $b \sim c$  then  $a \lesssim c$ ; and
- if  $a \lesssim b$  and  $a \sim c$  then  $c \lesssim b$ .

We also use the notation  $a < b$  indicating that  $a \lesssim b$  and  $a \not\sim b$ .

We first provide the following result; see Section 5.2.1 of [36]:

**PROPOSITION 6.** *Let  $\lesssim$  be a preorder over the set  $V$ , and  $\sim$  the equivalence relation on  $V$  as defined above. Let also  $Q$  be the set of all equivalence classes of  $V$  w.r.t.  $\sim$ . Then the relation  $\leq$  defined over  $Q$  by  $[[a] \leq [b] \iff a \lesssim b]$  is a partial order over  $Q$ .*

We say that a graph  $G = (V, E)$  admits a *valid preorder*  $\lesssim$  if, for nodes  $i$  and  $j$  of  $G$ , the following holds:

- if  $i \text{ --- } j$  then  $i \sim j$ ;
- if  $i \text{ --> } j$  then  $j < i$ ;
- if  $i \text{ <--> } j$  then  $i$  and  $j$  are incomparable.

The global interpretation of a valid preorder on graphs is as follows:

**PROPOSITION 7.** *Let  $\lesssim$  be a valid preorder for a graph  $G$ . It then holds for nodes  $i$  and  $j$  of  $G$  that if  $i \in \text{ant}(j)$  then  $j \lesssim i$ ; in particular,*

1. *if there is a semi-directed path from  $j$  to  $i$  then  $j < i$ ; and*
2. *if  $i$  and  $j$  are connected by a path consisting only of lines then  $i \sim j$ .*

**PROOF.** The proof follows from transitivity of both preorders and anterior paths.  $\square$

**COROLLARY 1.** *Let  $\lesssim$  be a valid preorder for a graph  $G$  and  $G_-$  its subgraph induced by lines. The equivalence classes of  $G$  based on  $\lesssim$  are the connected components of  $G_-$ . In addition,  $\leq$  defines a partial order over the connected components of  $G_-$  by [if  $i \text{ --> } j$ ,  $i \in \tau$  and  $j \in \delta$  then  $\tau > \delta$ ], for  $\tau$  and  $\delta$  two connected components of  $G_-$ .*

**PROOF.** The results follows from Propositions 7 and 6.  $\square$

For example, consider the graph in Fig. 1. The graph admits a valid preorder with the preorder provided over the nodes, or equivalently over the connected components of  $G_-$  (with the abuse of the notation for numbering). The notation implies  $2 > 1$  and  $2' > 1'$ , but  $a, b'$ , and  $c''$ ,  $a, b, c \in \{1, 2\}$ , are not comparable.

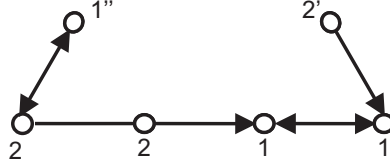


FIG 1. A preordered graph.

There may be many different preorders that are valid for a graph. If there is a valid preordering and we expect the other direction of Proposition 7 to hold then we obtain a unique preordering for the graph: Given a graph  $G$ , if  $i \notin \text{ant}(j)$  and  $j \notin \text{ant}(i)$  then set  $i$  and  $j$  to be incomparable. Otherwise, let  $i \sim j$  when  $i \text{ --- } j$ , and  $i > j$  when  $i \rightarrow j$ . It is easy to see that this, in fact, is a preorder for the nodes of  $G$ . We call this preordering the *minimal preorder* for  $G$  since it gives the fewest possible comparable pairs of nodes. For example, the preordering in the graph of Fig. 1 is minimal. In this paper, we mostly deal with this type of preorderings for graphs. It is easy to observe the following:

PROPOSITION 8. *If  $G$  is anterial then the minimal preorder for  $G$  is a valid preorder for  $G$ .*

In fact, in general, we have the following:

PROPOSITION 9. *A graph admits a valid preorder if and only if it is anterial.*

PROOF. If a graph  $G$  admits a valid preorder then by Proposition 7, there cannot be a directed cycle, nor can there be an arc with one endpoint anterior of the other. The converse is Proposition 8.  $\square$

Therefore, in addition to CMGs, SGs (and ADMGs) do not admit a valid preorder, but ancestral graphs do. However, notice that for every CMG, there exists a Markov equivalent AnG (see [34]); and the same for SGs and AG. Hence, the question of whether there is a CMG that is faithful to  $\mathcal{J}$  is the same as whether there is an AnG that is faithful to  $\mathcal{J}$ ; and the same holds for SG and AG.

As discussed above, there is a one-to-one correspondence between anterial graphs and the minimal preorder for the anterial graphs. When the skeleton of the concerned graph is known, we have the following trivial result:

PROPOSITION 10. *Given the skeleton of a graph  $G$  and a preorder  $\lesssim$  of its nodes,  $G$  can be constructed uniquely by directing the edges of  $G$ . In addition,  $\lesssim$  is a valid preorder for  $G$ .*

Let  $\mathcal{J}$  be defined over  $V$ . We introduce preordering in this paper in order to stress that, in principle, a preordering of members of  $V$  (which are possibly random variables) is defined irrespective of graphs to which they are Markov or faithful. However, for our purposes, knowing the skeleton of such graphs, one can directly discover the directions of the edges of the graph, which correspond to the minimal preordering for the graph, rather than working directly on the set  $V$ .

Suppose that there exists a preorder  $\lesssim$  over  $V$ . Proposition 10 implies that we can direct the edges of  $\text{sk}(\mathcal{J})$  based on  $\lesssim$  in order to define a *dependence graph*  $G(\mathcal{J}, \lesssim)$  induced by  $\mathcal{J}$  and  $\lesssim$ . Notice that, by Proposition 9,  $G(\mathcal{J}, \lesssim)$  is antierial. Similarly, for  $P$ , one can define  $G(P, \lesssim) = G(\mathcal{J}(P), \lesssim)$ .

As mentioned, we are only interested in preorderings that are minimal preorderings of graphs. Given  $\mathcal{J}$ , we define a preordering  $\lesssim$  to be  *$\mathcal{J}$ -compatible* if  $\lesssim$  is the minimal preorder for  $G(\mathcal{J}, \lesssim)$ . Similarly, one can define a  *$P$ -compatible preorder*.

4.2. *Ordered upward- and downward-stabilities.* We now exploit the preordering for independence models in order to define two other properties in addition to the seven properties defined in Section 2.3 (namely singleton-transitive compositional graphoid axioms). We say that an independence model  $\mathcal{J}$  over the set  $V$  respectively satisfies *ordered upward-* and *downward-stability* w.r.t. a preorder  $\lesssim$  of  $V$  if the following hold:

8. if  $\langle i, j | C \rangle \in \mathcal{J}$  then  $\langle i, j | C \cup \{k\} \rangle \in \mathcal{J}$  for every  $k \in V \setminus \{i, j\}$  such that  $l \lesssim k$  for some  $l \in \{i, j\}$  or  $l \sim k$  for some  $l \in C$  (ordered upward-stability);
9. if  $\langle i, j | C \rangle \in \mathcal{J}$  then  $\langle i, j | C \setminus \{k\} \rangle \in \mathcal{J}$  for every  $k \in V \setminus \{i, j\}$  such that  $l \not\lesssim k$  for every  $l \in \{i, j\}$  and  $l \not\prec k$  for every  $l \in C \setminus \{k\}$  (ordered downward-stability).

Ordered upward-stability is a generalization of a modification of *upward stability* defined in [9] for undirected graphs, with singletons instead of node subsets.

PROPOSITION 11. *Let  $G$  be an antierial graph and  $\lesssim$  the minimal preorder for  $G$ . Then  $\mathcal{J}(G)$  satisfies ordered upward- and downward-stability w.r.t.  $\lesssim$ .*

PROOF. First we prove that  $\mathcal{J}(G)$  satisfies ordered upward-stability w.r.t. the minimal preorder: If  $k \in C$ , the result is obvious, thus suppose that  $k \notin C$ . Suppose that there is a connecting walk  $\pi$  between  $i$  and  $j$  given  $C \cup \{k\}$ . If  $k$  is not on  $\pi$  then we are done; otherwise,  $k$  is on a collider section with no other node in  $C$  on  $\pi$ . Notice that  $l \lesssim k$  means here that  $k \in \text{ant}(l)$ , and  $l \sim k$  that  $k$  and  $l$  are connected by a path consisting of lines.

If  $l \in C$  then denote the path of lines between  $k$  and  $l$  by  $\omega$ . The walk consisting of the subwalk of  $\pi$  from  $i$  to  $k$ ,  $\omega$ ,  $\omega$  in the reverse direction (from  $l$  to  $k$ ), and the subwalk of  $\pi$  from  $k$  to  $j$  is connecting given  $C$ . This is because on this walk  $k$  and  $l$  are in the same collider section with  $l$  in  $C$ .

Now denote an anterior path from  $k$  to  $l \in \{i, j\}$  by  $\varpi$ . Without loss of generality we can assume that  $l = j$ . We now consider two cases: 1) If  $\varpi$  has no nodes in  $C$  then the walk containing the subwalk of  $\pi$  between  $i$  and  $k$  and  $\varpi$  is connecting given  $C$  between  $i$  and  $j$  since  $k$  and all nodes on  $\varpi$  are on non-collider sections on this walk. 2) If  $\varpi$  has a node in  $C$  then consider the one closest to  $k$  and call it  $h$ . Then the walk consisting of the subwalk of  $\pi$  from  $i$  to  $k$ , the subpath of  $\varpi$  from  $k$  to  $h$ , the reverse of the subpath from  $h$  to  $k$ , and the subwalk of  $\pi$  from  $k$  to  $j$  is connecting given  $C$  as proven as follows: 1) if there is an arrow on  $\varpi$  then  $k$  is on a non-collider section in both instances in which it appears on this walk, and  $h$  is on a collider section and in  $C$ . 2) if  $\varpi$  only consists of lines then  $k$  and  $h$  are in the same collider section with a node ( $h$ ) in  $C$ .

Now we prove that  $\mathcal{J}(G)$  satisfies ordered downward-stability: If  $k \notin C$  then the result is obvious, thus suppose that  $k \in C$ . Suppose that there is a connecting walk  $\pi$  between  $i$  and  $j$  given  $C \setminus \{k\}$ . If  $k$  is not on  $\pi$  then we are done. Otherwise, first it is clear that since  $\pi$  is connecting given  $C \setminus \{k\}$ ,  $k$  cannot be the only node on a collider section that is in  $C$ .

If  $k$  is on the same section as another member of  $C$  then  $\pi$  is clearly connecting given  $C$ . The only case that is left is when  $k$  is on a non-collider section on  $\pi$ ; but this is impossible: there is no arrowhead at the section containing  $k$  from one side on  $\pi$ , say from the  $j$  side. By moving towards  $j$  from  $k$ , it can be implied that  $k$  is either an anterior of a node in a collider section that is in  $C$ , or an anterior of  $j$ , but both are impossible since the former implies  $l < k$  for  $l \in C$ , and the latter implies  $j \lesssim k$ .  $\square$

## 5. Characterization of graphical independence models.

5.1. *Sufficient conditions for minimal Markovness.* Here we provide sufficient conditions for an independence model to be minimally Markov to a graph. First, we have the following trivial result:



PROPOSITION 12. *If  $\mathcal{J}$  is minimally Markov to an anterial graph  $G$  then  $G = G(\mathcal{J}, \lesssim)$  for the minimal preorder  $\lesssim$  for  $G$ .*

By Proposition 5, the above statement also holds when  $\mathcal{J}$  and  $G$  are faithful.

PROPOSITION 13. *Suppose that there exists a  $\mathcal{J}$ -compatible preorder  $\lesssim$  over  $V$  w.r.t. which  $\mathcal{J}$  satisfies ordered downward- and upward-stability. It then holds that  $\mathcal{J}$  is pairwise Markov to  $G(\mathcal{J}, \lesssim)$ .*

PROOF. Suppose that there are non-adjacent nodes  $i$  and  $j$  in  $G(\mathcal{J}, \lesssim)$ . These are non-adjacent in  $\text{sk}(\mathcal{J})$  too. By definition of  $\text{sk}(\mathcal{J})$ , it holds that  $\langle i, j \mid C \rangle \in \mathcal{J}$  for some  $C \subseteq V \setminus \{i, j\}$ .

Since  $\mathcal{J}$  satisfies ordered downward-stability and since the preorder is  $\mathcal{J}$ -compatible, as long as a node  $k \in C$  is not an anterior of  $\{i, j\}$  or there is no semi-directed path from it to  $C \setminus \{k\}$ , it can be removed from the conditioning set, i.e., we have that  $\langle i, j \mid C \setminus k \rangle \in \mathcal{J}$ . Since  $G(\mathcal{J}, \lesssim)$  is acyclic, as long as  $\text{ant}(\{i, j\}) \subset C$ , such a  $k$  exists. Now consider a node  $k' \in C \setminus \{k\}$  that is not an anterior of  $\{i, j\}$  or there is no semi-directed path from it to  $C \setminus \{k, k'\}$  and apply downward-stability again. By an inductive argument we imply that  $i \perp\!\!\!\perp j \mid C \cap \text{ant}(\{i, j\})$ . Now since  $P$  satisfies ordered upward-stability, nodes outside  $C$  that are in  $\text{ant}(\{i, j\})$  can be added to the conditioning set. Hence, it holds that  $\langle i, j \mid \text{ant}(\{i, j\}) \setminus \{i, j\} \rangle \in \mathcal{J}$ . This completes the proof.  $\square$

THEOREM 1. *Suppose, for an independence model  $\mathcal{J}$  over  $V$ , that*

1.  *$\mathcal{J}$  is a compositional graphoid; and*
2. *there exists a  $\mathcal{J}$ -compatible preorder  $\lesssim$  over  $V$  w.r.t. which  $\mathcal{J}$  satisfies ordered downward- and upward-stability.*

*It then holds that  $\mathcal{J}$  is minimally Markov to  $G(\mathcal{J}, \lesssim)$ .*

PROOF. The proof follows from Proposition 13, the fact that  $\text{sk}(\mathcal{J}) = \text{sk}(G(\mathcal{J}, \lesssim))$ , and Proposition 3.  $\square$

5.2. *Conditions for faithfulness.* In this section, we present the main result of this paper, which provides necessary and sufficient conditions for faithfulness of independence models (and probability distributions) and a (not necessarily known) graph in the general case. In Section 6, we specialize the result to the more well-known subclasses as corollaries.

PROPOSITION 14. *Let  $\mathcal{J}$  be an independence model over  $V$ . Suppose that  $\mathcal{J}$  is minimally Markov to an antierial graph  $G$ . It then holds that  $\mathcal{J}$  and  $G$  are faithful if and only if*

1.  $\mathcal{J}$  satisfies singleton-transitivity; and
2.  $\mathcal{J}$  satisfies ordered downward- and upward-stability w.r.t. the minimal preorder for  $G$ .

PROOF. Suppose that  $\mathcal{J}$  and  $G$  are faithful. This is equivalent to  $\mathcal{J} = \mathcal{J}(G)$ . By Proposition 2,  $\mathcal{J}$  satisfies singleton-transitivity, and by Proposition 11, the result follows.

Conversely, suppose that  $\mathcal{J}$  satisfies singleton-transitivity as well as ordered downward- and upward-stability w.r.t. the minimal preorder  $\lesssim$  for  $G$ . By Proposition 12,  $G = G(\mathcal{J}, \lesssim)$ . We show that  $\mathcal{J}$  and  $G$  are faithful. We need to show that if  $\langle A, B | C \rangle \in \mathcal{J}$  then  $A \perp B | C$  in  $G$ . Consider  $i \in A$  and  $j \in B$ . By decomposition, we have that  $\langle i, j | C \rangle \in \mathcal{J}$ , which implies that  $i$  and  $j$  are not adjacent in  $G$ .

If, for contradiction,  $C$  does not separate nodes  $i$  and  $j$  then there exists a connecting walk  $\pi = \langle i = i_1, i_2, \dots, i_r = j \rangle$ , on which all collider sections have a node in  $C$  and all non-collider sections are outside  $C$ . In addition, for every node  $i_q$ ,  $2 \leq q \leq r-1$ , on  $\pi$  define  $C_{i_q}$  as follows: If  $i_q \notin C$  then  $C_{i_q} = \emptyset$ ; if  $i_q \in C$  then  $C_{i_q}$  is the set of all  $l \in C \setminus \pi$  such that there is a semi-directed path from  $i_q$  to  $l$  but not from any  $i_p \in C$  to  $l$ ,  $p \in \{q+1, \dots, r-1\}$ . Let  $C_\pi = \bigcup_{q=2}^{r-1} C_{i_q}$ .

We prove by induction along nodes of  $\pi$  that for every  $q$ ,  $2 \leq q \leq r$ , it holds that  $\langle i_1, i_q | (\bigcup_{p=2}^{q-1} C_{i_p}) \cup [C \setminus (\{i_q, \dots, i_r\} \cup C_\pi)] \rangle \notin \mathcal{J}$ . The base for  $q = 2$  is trivial since by the construction of  $G(\mathcal{J}, \lesssim)$ , adjacent nodes are dependent given anything. Suppose now that the result holds for  $q$  and we prove it for  $q+1$ .

First, again by the construction of  $G(\mathcal{J}, \lesssim)$ , it holds that  $\langle i_q, i_{q+1} | (\bigcup_{p=2}^{q-1} C_{i_p}) \cup [C \setminus (\{i_q, \dots, i_r\} \cup C_\pi)] \rangle \notin \mathcal{J}$ . By singleton-transitivity, this and the induction hypothesis imply (a):  $\langle i_1, i_{q+1} | (\bigcup_{p=2}^{q-1} C_{i_p}) \cup [C \setminus (\{i_q, \dots, i_r\} \cup C_\pi)] \rangle \notin \mathcal{J}$ ; or (b):  $\langle i_1, i_{q+1} | \{i_q\} \cup (\bigcup_{p=2}^{q-1} C_{i_p}) \cup [C \setminus (\{i_q, \dots, i_r\} \cup C_\pi)] \rangle \notin \mathcal{J}$ . We now have two cases:

1) If  $i_q$  is on a collider section: The section containing  $i_q$  has a node in  $C$ . We again consider two subcases:

1.i) If  $i_q$  is in  $C$ : If (a) holds then first notice that  $i_q$  is not an anterior of  $\{i_1, i_{q+1}\}$  nor is there a semi-directed path from  $i_q$  to  $(\bigcup_{p=2}^{q-1} C_{i_p}) \cup [C \setminus (\{i_q, \dots, i_r\} \cup C_\pi)]$ . We apply downward-stability to (a) to obtain (b). If  $C_{i_q} \neq \emptyset$  then consider a highest order node  $l$  in  $C_{i_q}$ . Again by downward-stability, we obtain  $\langle i_1, i_{q+1} | \{l\} \cup (\bigcup_{p=2}^{q-1} C_{i_p}) \cup [C \setminus (\{i_{q+1}, \dots, i_r\} \cup C_\pi)] \rangle \notin$

$\mathcal{J}$ . By an inductive argument along the members of  $C_{i_q}$  (by, at each step, choosing a highest order node in  $C_{i_q}$  that has not yet been chosen), we eventually obtain  $\langle i_1, i_{q+1} \mid (\cup_{p=2}^q C_{i_p}) \cup [C \setminus (\{i_{q+1}, \dots, i_r\} \cup C_\pi)] \rangle \notin \mathcal{J}$ . Notice that  $i_q$  is in the conditioning set of this dependency.

1.ii) If  $i_q$  is not in  $C$ : If (b) holds then observe that  $i_q$  is in the same equivalence class as that of either  $i_{q+1}$  or a node  $i_s \in C$ ,  $s < q$ . Hence we can apply upward-stability to obtain (a). Since  $C_{i_2} = \emptyset$ , (a) is clearly the same as  $\langle i_1, i_{q+1} \mid (\cup_{p=2}^q C_{i_p}) \cup [C \setminus (\{i_{q+1}, \dots, i_r\} \cup C_\pi)] \rangle \notin \mathcal{J}$ . Notice that  $i_q$  is not in the conditioning set of this dependency.

2) If  $i_q$  is on a non-collider section: We have that  $i_q \notin C$ , and  $i_q \in \text{ant}(\{i_1, i_{q+1}\})$ . Hence, by upward-stability, from (b) we obtain (a). Again, since  $C_{i_q} = \emptyset$ , we obtain the result.

Therefore, we established that  $\langle i_1, i_q \mid (\cup_{p=2}^{q-1} C_{i_p}) \cup [C \setminus (\{i_q, \dots, i_r\} \cup C_\pi)] \rangle \notin \mathcal{J}$ , for all  $q$ , where depending on whether  $i_q \in C$  or not, the conditioning set contains or does not contain  $i_q$ . By letting  $q = r$ , we obtain  $\langle i_1, i_r \mid C \rangle \notin \mathcal{J}$ , a contradiction. Therefore, it holds that  $i \perp j \mid C$  in  $G$ . By the composition property for separation for graphs (Proposition 1), we obtain the result.  $\square$

We follow the arguments in the proof for the following example.

EXAMPLE 1. Consider the undirected (LWF) chain graph  $G$  in Fig. 2. Suppose that a probability distribution  $P$  is minimally Markov to  $G$ , and  $\mathcal{J}(P)$  satisfies singleton-transitivity and ordered downward- and upward-stability w.r.t. the minimal preorder for  $G$ . We have that  $l \not\perp j \mid l_2$ . Notice that this is a case where there is no path that connects  $l$  and  $j$ . In order to show that  $l \not\perp j \mid l_2$ , we start from  $l = l_1$  and move towards  $j = l_7$  via the connecting walk  $\langle l_1, l_2, l_3, l_4, l_5, l_6, l_7 \rangle$  given  $l_2$ .

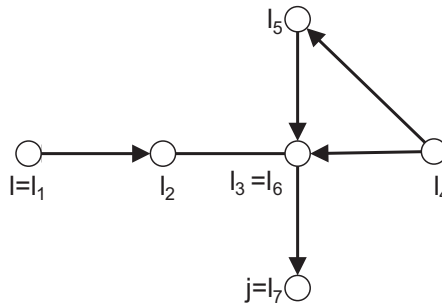


FIG 2. An undirected (LWF) chain graph.

We abbreviate singleton-transitivity, ordered upward-stability, and ordered downward stability by “st”, “ous”, and “ods” respectively.

$$\begin{array}{ccccccc}
l_1 \not\perp l_2 & l_1 \not\perp l_3 & l_4 \not\perp l_5 \mid l_2 & l_1 \not\perp l_5 \mid \{l_2, l_4\} & l_6 \not\perp l_7 \mid l_2 & l_1 \not\perp l_7 \mid \{l_2, l_6\} \\
\Downarrow \text{st or } \Downarrow \text{ods} & & \Downarrow \text{st or } \Downarrow \text{ous} & & \Downarrow \text{st or } \Downarrow \text{ous} & \\
l_2 \not\perp l_3 & l_1 \not\perp l_3 \mid l_2 & l_1 \not\perp l_4 \mid l_2 & l_1 \not\perp l_5 \mid l_2 & l_1 \not\perp l_6 \mid l_2 & l_1 \not\perp l_7 \mid l_2. \\
\Downarrow \text{st or } \Uparrow \text{ous} & & \Downarrow \text{st or } \Uparrow \text{ous} & & \Downarrow \text{st or } \Uparrow \text{ous} & \\
l_3 \not\perp l_4 \mid l_2 & l_1 \not\perp l_3 \mid \{l_2, l_3\} & l_5 \not\perp l_6 \mid l_2 & l_1 \not\perp l_6 \mid \{l_2, l_5\} & & 
\end{array}$$

The following example shows how singleton-transitivity and ordered stabilities are necessary for faithfulness.

EXAMPLE 2. Consider the DAG  $G$  in Fig. 3. If an independence model  $\mathcal{I}$  is minimally Markov to  $G$  then we must have  $\langle h, k \mid \{j, l\} \rangle \in \mathcal{I}$  and  $\langle j, l \mid k \rangle \in \mathcal{I}$ . Now, in violation of faithfulness, if only  $\langle h, k \mid \emptyset \rangle \in \mathcal{I}$  in addition to these holds then upward-stability is violated since for  $j > h$ ,  $\langle h, k \mid j \rangle \in \mathcal{I}$  must hold. If, in violation of faithfulness, only  $\langle h, k \mid j \rangle \in \mathcal{I}$  in addition to the original statements holds then singleton-transitivity is violated since  $\langle h, k \mid j \rangle \in \mathcal{I}$  and  $\langle h, k \mid \{j, l\} \rangle \in \mathcal{I}$  must imply either  $\langle h, l \mid j \rangle \in \mathcal{I}$  or  $\langle k, l \mid j \rangle \in \mathcal{I}$ , which are both impossible due to minimality since  $h, l$  and  $k, l$  are adjacent in  $G$ .

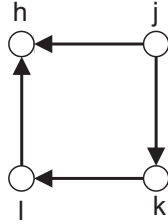


FIG 3. A DAG.

Without assuming the Markov assumption we have the following:

THEOREM 2. Let  $\mathcal{I}$  be an independence model defined over  $V$ . It then holds that  $\mathcal{I}$  is graphical if and only if

1.  $\mathcal{I}$  is a singleton-transitive compositional graphoid; and
2. there exists a  $\mathcal{I}$ -compatible preorder  $\lesssim$  over  $V$  w.r.t. which  $\mathcal{I}$  satisfies ordered downward- and upward-stability.

In addition, if 1 and 2 hold then  $\mathcal{I}$  is faithful to  $G(\mathcal{I}, \lesssim)$ .

PROOF. ( $\Rightarrow$ ) follows from Propositions 1, 2, and 11, and the fact that there is always an antierial graph that is Markov equivalent to a given CMG.

To prove  $(\Leftarrow)$  and the last statement of the theorem, consider  $G(\mathcal{J}, \lesssim)$ . Theorem 1 implies that  $\mathcal{J}$  is minimally Markov to  $G(\mathcal{J}, \lesssim)$ . This together with the  $\mathcal{J}$ -compatibility of the preorder, using Proposition 14, implies that  $\mathcal{J}$  and  $G(\mathcal{J}, \lesssim)$  are faithful.  $\square$

Hence, for probability distributions, we have the following characterization:

**COROLLARY 2.** *Let  $P$  be a probability distribution defined over  $\{X_\alpha\}_{\alpha \in V}$ . It then holds that  $P$  is graphical if and only if*

1.  $\mathcal{J}(P)$  satisfies intersection, composition, and singleton-transitivity; and
2. there exists a  $P$ -compatible preorder  $\lesssim$  over  $V$  w.r.t. which  $\mathcal{J}(P)$  satisfies ordered downward- and upward-stability.

*In addition, if 1 and 2 hold then  $P$  is faithful to  $G(P, \lesssim)$ .*

In fact, we have shown in Theorem 2 that if  $\mathcal{J}$  is graphical then for every  $\mathcal{J}$ -compatible preorder w.r.t. which  $\mathcal{J}$  satisfies ordered downward- and upward-stability,  $G(\mathcal{J}, \lesssim)$  is a graph that is faithful to  $\mathcal{J}$ . A consequence of Proposition 12 implies that such graphs (i.e.  $G = G(\mathcal{J}, \lesssim)$  for some preorder  $\lesssim$  w.r.t. which  $\mathcal{J}$  satisfies ordered downward- and upward-stability) constitute the set of all graphs that are faithful to  $\mathcal{J}$ , i.e. a Markov equivalence class of graphs whose members are faithful to  $\mathcal{J}$ . Every member of the equivalence class corresponds to a different preorder w.r.t. which  $\mathcal{J}$  satisfies ordered downward- and upward-stability.

Of course, if the goal is to verify whether  $\mathcal{J}$  and a given  $G$  are faithful then we can use the following corollary:

**COROLLARY 3.** *Let  $\mathcal{J}$  be an independence model defined over  $V$ , and  $G$  an ancestral graph with node set  $V$ . It then holds that if*

1.  $\mathcal{J}$  is a singleton-transitive compositional graphoid; and
2.  $\mathcal{J}$  satisfies ordered downward- and upward-stability w.r.t. the minimal preorder for  $G$ ,

*then  $\mathcal{J}$  and  $G$  are faithful.*

It is also important to note that singleton-transitivity and ordered upward or downward-stability are necessary in the sense that they are not implied by one another, although there are different ways to axiomatize singleton-transitive compositional graphoids that satisfy ordered upward- and downward-stability.

**6. Specialization to subclasses.** The definitions and results presented in this paper can be specialized to any subclass of CMGs including ancestral graphs and LWF chain graphs. However, here we only focus on the three of the most used classes of undirected (concentration), bidirected (covariance), and directed acyclic graphs. We also mostly focus on probabilistic independence models in this section.

6.1. *Specialization to undirected and bidirected graphs.* For connected UGs, every two nodes are in the same equivalence class, and for BGs every two nodes are incomparable; hence, the minimal (pre)ordering is trivial and uninteresting in these cases. More precisely, ordered upward- and downward-stabilities can be specialized for these two types of trivial preordering:

- 8a. if  $\langle i, j | C \rangle \in \mathcal{J}$  then  $\langle i, j | C \cup \{k\} \rangle \in \mathcal{J}$  for every  $k \in V \setminus \{i, j\}$  (*upward-stability*);
- 9a. if  $\langle i, j | C \rangle \in \mathcal{J}$  then  $\langle i, j | C \setminus \{k\} \rangle \in \mathcal{J}$  for every  $k \in V \setminus \{i, j\}$  (*downward-stability*).

PROPOSITION 15. *An independence model  $\mathcal{J}$  satisfies ordered upward- and downward-stability (8) and (9) w.r.t. the minimal preorder if and only if,*

- 1. *for connected undirected graphs, it satisfies upward-stability (8a);*
- 2. *for bidirected graphs, it satisfies downward-stability (9a).*

*In addition, if, for undirected graphs,  $\mathcal{J}$  satisfies (8a) then it satisfies (8).*

PROOF. The proof of 1 and 2 are straightforward since in 1 all nodes are in the same equivalence class, and in 2 all nodes are incomparable. The second part is trivial.  $\square$

Analogous to Proposition 11, it is also straightforward to show the following:

PROPOSITION 16. *It holds that*

- 1. *for an undirected graph  $G$ ,  $\mathcal{J}(G)$  satisfies upward-stability;*
- 2. *for a bidirected graph  $G$ ,  $\mathcal{J}(G)$  satisfies downward-stability.*

For UGs and BGs, denote the unique and trivial valid preorders by  $\lesssim_u^*$  and  $\lesssim_b^*$  respectively. For UGs,  $G_u^*(P) := G(P, \lesssim_u^*) = \text{sk}(P)$ , whereas for BGs,  $G_b^*(P) := G(P, \lesssim_b^*)$  is  $\text{sk}(P)$  with all edges directed to be bidirected. Another way to construct induced UGs and BGs by  $P$  is to construct them based on their corresponding pairwise Markov property, i.e. connect every

pair of nodes  $i$  and  $j$  if  $i \not\perp j \mid V \setminus \{i, j\}$  for UGs, and if  $i \not\perp j$  for BGs. Denote these graphs by  $G_u(P)$  and  $G_b(P)$  respectively. We then have the following:

PROPOSITION 17. *It holds that,*

1. *for undirected graphs,  $G_u^*(P) = G_u(P)$  if  $\mathcal{J}(P)$  satisfies upward-stability;*
2. *for bidirected graphs,  $G_b^*(P) = G_b(P)$  if  $\mathcal{J}(P)$  satisfies downward-stability.*

For upward- and downward-stability, we also have the following observation:

LEMMA 1. *For a probability distribution  $P$  the following holds:*

1. *If  $\mathcal{J}(P)$  satisfies upward-stability then  $\mathcal{J}(P)$  satisfies composition.*
2. *If  $\mathcal{J}(P)$  satisfies downward-stability then  $\mathcal{J}(P)$  satisfies intersection.*

PROOF. 1. Suppose that  $A \perp\!\!\!\perp B \mid C$  and  $A \perp\!\!\!\perp D \mid C$ . By decomposition, it holds, for  $i \in A$ ,  $j \in B$ , and  $k \in D$ , that  $i \perp\!\!\!\perp j \mid C$  and  $i \perp\!\!\!\perp k \mid C$ . By upward-stability, we obtain  $i \perp\!\!\!\perp j \mid C \cup \{k\}$ . Contraction implies  $i \perp\!\!\!\perp \{j, k\} \mid C$ . By an inductive argument, we obtain the result.

2. Suppose that  $A \perp\!\!\!\perp B \mid C \cup D$  and  $A \perp\!\!\!\perp D \mid C \cup B$ . By decomposition, it holds, for  $i \in A$ ,  $j \in B$ , and  $k \in D$ , that  $i \perp\!\!\!\perp j \mid C \cup \{k\}$  and  $i \perp\!\!\!\perp k \mid C \cup \{j\}$ . By downward-stability, we obtain  $i \perp\!\!\!\perp k \mid C$ . Contraction implies  $i \perp\!\!\!\perp \{j, k\} \mid C$ . By an inductive argument, we obtain the result.  $\square$

LEMMA 2. *Let  $P$  be a probability distribution, and suppose that  $\mathcal{J}(P)$  satisfies intersection. It then holds that  $\mathcal{J}(P)$  satisfies ordered downward-stability (9) w.r.t. the minimal preorder for  $G_u(P)$ .*

PROOF. By definition, we know that  $P$  is pairwise Markov to  $G_u(P)$ . It is known that under intersection, this implies that  $P$  is Markov to  $G_u(P)$ ; see [27]. Now suppose that  $i \perp\!\!\!\perp j \mid C$ . The only applicable  $k$  in the definition of (9) is not in the same connected component as that of  $i$  or  $j$ . Hence, since  $i \perp\!\!\!\perp j \mid C \setminus \{k\}$ , by the global Markov property, it holds that  $i \perp\!\!\!\perp j \mid C \setminus \{k\}$ .  $\square$

As a consequence of Theorem 1, we have the following for UGs and BGs:

COROLLARY 4. *It holds that*

1. *if  $\mathcal{J}(P)$  satisfies intersection and upward-stability then  $P$  is minimally Markov to the undirected graph  $G_u(P)$ ;*
2. *if  $\mathcal{J}(P)$  satisfies composition and downward-stability then  $P$  is minimally Markov to the bidirected graph  $G_b(P)$ .*



PROOF. The proof follows from Propositions 17 and 15, and Lemmas 1 and 2.  $\square$

For other classes of graphs, the preordering (or ordering) of nodes (and consequently random variables) is necessary since the conditioning set  $C(i, j)$  for the pairwise Markov property depends on the preordering of the variables. This implies that in general one cannot construct more than the skeleton of graphs based on the pairwise Markov property if there is no given preordering.

Proposition 14 specializes to the following two corollaries:

COROLLARY 5. *Suppose that a distribution  $P$  is Markov to  $G_u(P)$ . Then  $P$  and  $G_u(P)$  are faithful if and only if  $\mathcal{J}(P)$  satisfies singleton-transitivity and upward-stability.*

PROOF. The proof follows from Propositions 15, 16, and 17, and Lemma 2.  $\square$

Similarly, for BGs, we have the following:

COROLLARY 6. *Suppose that a distribution  $P$  is Markov to  $G_b(P)$ . Then  $P$  and  $G_b(P)$  are faithful if and only if  $\mathcal{J}(P)$  satisfies singleton-transitivity and downward-stability.*

The main conditions for a distribution to be faithful to a UG or BG (consequence of Theorem 2) are as follows:

COROLLARY 7. *Let  $P$  be a probability distribution defined over  $\{X_\alpha\}_{\alpha \in V}$ . It then holds that  $P$  is faithful to an undirected graph if and only if  $\mathcal{J}(P)$  satisfies intersection, singleton-transitivity, and upward-stability. In addition, if the conditions hold then  $P$  is faithful to  $G_u(P)$ .*

PROOF. The proof follows from Propositions 15, 16, and 17, and Lemmas 1 and 2.  $\square$

COROLLARY 8. *Let  $P$  be a probability distribution defined over  $\{X_\alpha\}_{\alpha \in V}$ . It then holds that  $P$  is faithful to a bidirected graph if and only if  $\mathcal{J}(P)$  satisfies composition, singleton-transitivity, and downward-stability. In addition, if the conditions hold then  $P$  is faithful to  $G_b(P)$ .*

6.2. *Specialization to directed acyclic graphs.* For DAGs, the valid pre-order specializes to a valid (partial) order since there are no lines in the graph: If  $i \rightarrow j$  then  $i > j$ . In the minimal order for a DAG, two nodes are incomparable if and only if neither is an ancestor of the other. Therefore, we define ordered stabilities w.r.t. ordering  $\leq$  for DAGs:

- 8b. if  $\langle i, j | C \rangle \in \mathcal{J}$  then  $\langle i, j | C \cup \{k\} \rangle \in \mathcal{J}$  for every  $k \in V \setminus \{i, j\}$  such that  $l < k$  for some  $l \in \{i, j\}$  (*ordered upward-stability*);
- 9b. if  $\langle i, j | C \rangle \in \mathcal{J}$  then  $\langle i, j | C \setminus \{k\} \rangle \in \mathcal{J}$  for every  $k \in V \setminus \{i, j\}$  such that  $l \not\prec k$  for every  $l \in \{i, j\} \cup C \setminus \{k\}$  (*ordered downward-stability*).

As a consequence of Proposition 11, we have the following:

**COROLLARY 9.** *For a directed acyclic graph  $G$ , the induced independence model by  $d$ -separation  $\mathcal{J}(G)$  satisfies ordered upward- and downward-stability (8b) and (9b) w.r.t. the minimal order for  $G$ .*

As a consequence of Theorem 1, and for  $G(P, \leq)$ , the DAG obtained by directing the edges of  $\text{sk}(P)$  based on the order  $\leq$ , we have the following:

**COROLLARY 10.** *Suppose, for a probability distribution  $P$  over  $\{X_\alpha\}_{\alpha \in V}$ , that*

- 1.  $\mathcal{J}(P)$  satisfies intersection and composition properties; and
- 2. there exists a  $P$ -compatible order  $\leq$  over  $V$  w.r.t. which  $\mathcal{J}(P)$  satisfies ordered downward- and upward-stability, (8b) and (9b).

*Then  $P$  is minimally Markov to the directed acyclic graph  $G(P, \leq)$ .*

The main conditions for a distribution to be faithful to a DAG (consequence of Theorem 2) is as follows:

**COROLLARY 11.** *Let  $P$  be a probability distribution defined over  $\{X_\alpha\}_{\alpha \in V}$ . It then holds that  $P$  is faithful to a directed acyclic graph  $G$  if and only if*

- 1.  $\mathcal{J}(P)$  satisfies intersection, composition, and singleton-transitivity; and
- 2. there exists a  $P$ -compatible order  $\leq$  over  $V$  w.r.t. which  $\mathcal{J}(P)$  satisfies ordered downward- and upward-stability, (8b) and (9b).

In addition, if the conditions hold then  $P$  is faithful to  $G(P, \leq)$ .

**7. Comparison to the results in the literature.** All results in the literature concerning faithfulness of certain probability distributions and certain types of graphs can be considered corollaries (or examples) of Theorem

2. Here we verify this for some of the more well-known or interesting results. For brevity, we leave out some interesting results such as faithfulness of Gaussian and discrete distributions and LWF chain graphs; see [25] and [24; 42] respectively.

*Faithfulness of  $\text{MTP}_2$  distributions and undirected graphs.* Multivariate totally positive of order 2 ( $\text{MTP}_2$ ) distributions [13] are distributions whose density  $f$  is  $\text{MTP}_2$ , that is  $f(x)f(y) \leq f(x \wedge y)f(x \vee y)$ , for all possible vectors  $x, y$ , where  $x \wedge y = (\min(x_1, y_1), \dots, \min(x_m, y_m))$  and  $x \vee y = (\max(x_1, y_1), \dots, \max(x_m, y_m))$ . In [9], it was shown that when an  $\text{MTP}_2$  distribution  $P$  is a graphoid, it holds that  $P$  and what we denote here by  $G_u(P)$  are faithful. It is known that  $\text{MTP}_2$  distributions satisfy both singleton-transitivity and upward-stability, but not necessarily the intersection property; see [9]. Hence, this result is a direct implication of Corollary 7.

*Faithfulness of Gaussian distributions and undirected graphs.* [21] provides regular Gaussian distributions that are faithful to any undirected graph; see Corollary 3 of the mentioned paper. In order to do so, given an undirected graph  $G = (V, E)$ , they provide the following matrix  $A^{G, \epsilon}$ , which acts as the covariance matrix of a Gaussian distribution that is faithful to  $G$ :

$$A^{G, \epsilon} = \begin{cases} 1 & \text{if } i = j, \\ \epsilon & \text{if } i \neq j \text{ and } i \text{ and } j \text{ are adjacent in } G, \\ 0 & \text{otherwise,} \end{cases}$$

where  $i, j \in V$ .

It is known that, for sufficiently small  $\epsilon > 0$ , the corresponding Gaussian distribution is regular and satisfies singleton-transitivity and intersection as required in the conditions of Corollary 7. The only thing that is left to be shown is that it satisfies upward-stability. It is straightforward to show that for sufficiently small  $\epsilon > 0$ , the inverse of  $A^{G, \epsilon}$ , which is the concentration matrix of the Gaussian distribution, is an  $M$ -matrix, i.e., all its on-diagonal elements are positive and all its off-diagonal elements are non-positive. This is equivalent to the Gaussian distribution being  $\text{MTP}_2$ ; see [14]. Hence, upward-stability is satisfied as discussed above.

*Unfaithfulness of certain Gaussian distributions and undirected graphs and DAGs.* There are many examples in the literature for, and in fact it is easy to construct examples of, probability distributions that are not faithful to undirected graphs; for one of the first observations of this behaviour in a data set, see Example 4 of [5]. As a constructed example, let us discuss Example 1 of [37], which provided the Gaussian distribution for  $X = (X_1, X_2, X_3, X_4)$

with zero mean and positive definite covariance matrix

$$\Sigma = \begin{bmatrix} 3 & 2 & 1 & 2 \\ 2 & 4 & 2 & 1 \\ 1 & 2 & 7 & 1 \\ 2 & 1 & 1 & 6 \end{bmatrix}.$$

It can be seen that the concentration matrix  $\Sigma^{-1}$  has no zero entries, but  $X_1 \perp\!\!\!\perp X_3 \mid X_2$  since  $\Sigma_{13} = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{23}$ . Hence, the distribution and its complete skeleton are not faithful. This is clearly a consequence of the violation of upward-stability.

For DAGs too, it is easy to construct or observe examples where faithfulness is violated. One of the simplest of such examples, is what is related to the so-called “transitivity of causation” in philosophical causality; see, e.g. [30], where the goal is to show how the PC algorithm [38] errs. Consider the graph  $a \rightarrow b \rightarrow c$ , where  $a \perp\!\!\!\perp c$  and  $a \perp\!\!\!\perp c \mid b$ . This, in our language, is a clear violation of singleton-transitivity since the graph implies that  $a, b$  and  $b, c$  are always dependent.

*Faithfulness of Gaussian and discrete distributions and DAGs.* It was shown in [22] that, in certain measure-theoretic senses, almost all the regular Gaussian and discrete distributions that are Markov to a given DAG are faithful to it. It is known, for DAGs, that the global Markov property is equivalent to the *factorization property* (see [17]), which holds if the joint density can factorize as follows:

$$f_V(x) = \prod_{v \in V} f_V(x_v \mid x_{\text{pa}(v)}).$$

Since Gaussian and discrete distributions satisfy singleton-transitivity, in order to prove the result, one needs to show that ordered upward- and downward-stabilities, (8b) and (9b), are satisfied for almost all distributions that factorize w.r.t. the DAG. We refrain from showing this explicitly here due to technicality of the proof, but the general idea is similar to the proofs in Appendix B of [22]: For (9b), we have that the probabilistic conditional independence holds for the marginal probability over  $(i, j, C, k)$ , and we need to exploit the factorization to show that the same conditional independence formula holds for the marginal probability over  $(i, j, C)$  when  $k \notin \text{an}(\{i, j\} \cup C)$ . In this case the corresponding resulting polynomials are much simpler than those of [22] since one needs to consider extra marginalization over only one variable  $k$ . For (8b), we will need to prove the other direction in a similar fashion.

*Faithfulness of Gaussian distributions and ancestral Graphs.* The main goal of [32] is to capture the conditional independence structure of a DAG model after marginalization and conditioning; see also [33]. This was done by introducing the class of ancestral graphs, and it was shown that every ancestral graph is probabilistic. In fact, there is a Gaussian distribution faithful to any ancestral graph, which is indeed the Gaussian distribution faithful to a DAG after marginalization and conditioning; see Theorem 7.5 of this paper.

However, by using our results, without going through the standard theory of ancestral graphs, it is possible to show that distributions that are faithful to a DAG are graphical after marginalization and conditioning: In order to do so, we use the notation in [33], for an independence model  $\mathcal{J}$ , to define the *independence model after marginalization over  $M$  and conditioning on  $C$* , denoted by  $\alpha(\mathcal{J}; M, C)$ , for disjoint subsets  $M$  and  $C$  of  $V$ :

$$\alpha(\mathcal{J}; M, C) = \{\langle A, B \mid D \rangle : \langle A, B \mid D \cup C \rangle \in \mathcal{J} \text{ and } (A \cup B \cup D) \cap (M \cup C) = \emptyset\}.$$

First, we show that  $\alpha(\mathcal{J}; M, C)$  satisfies condition 1 of Theorem 2.

PROPOSITION 18. *If  $\mathcal{J}$  satisfies singleton-transitivity, intersection, and composition then so does  $\alpha(\mathcal{J}; M, C)$ .*

PROOF. For composition, assume that  $\langle A, B \mid C' \rangle \in \alpha(\mathcal{J}; M, C)$  and  $\langle A, D \mid C' \rangle \in \alpha(\mathcal{J}; M, C)$ . By definition, we have  $\langle A, B \mid C' \cup C \rangle \in \mathcal{J}$  and  $\langle A, D \mid C' \cup C \rangle \in \mathcal{J}$ . Composition for  $\mathcal{J}$  implies  $\langle A, B \cup D \mid C' \cup C \rangle \in \mathcal{J}$ . Therefore,  $\langle A, B \cup D \mid C' \rangle \in \alpha(\mathcal{J}; M, C)$ . Proof of the others is similar to this one and as straightforward.  $\square$

Denote by  $\alpha(G; M, C)$ , the generated ancestral graph from the DAG  $G$  and  $M$  and  $C$  two disjoint subsets of its node set that are marginalized over and conditioned on respectively. Here we do not go through the generating process of ancestral graphs from DAGs; for more details, see [32; 33]. Without explicitly using the theory of stability under marginalization and conditioning for ancestral graphs, we show the following.

PROPOSITION 19. *If  $\mathcal{J}$  and a DAG  $G$  are faithful then  $\alpha(\mathcal{J}; M, C)$  and the ancestral graph  $\alpha(G; M, C)$  are faithful.*

PROOF. By Proposition 18, we only need to show that w.r.t. the ordering associated with  $\alpha(G; M, C)$ , condition 2 of Theorem 2 is also satisfied by  $\alpha(\mathcal{J}; M, C)$ .

To prove ordered upward-stability, assume that  $\langle i, j \mid C' \cup \{k\} \rangle \notin \alpha(\mathcal{J}; M, C)$ , where  $k$  is an anterior of a node  $l$ ,  $l \in \{i, j\}$  or connected by lines to  $l$ ,  $l \in C'$

in  $H = \alpha(G; M, C)$ . We have that  $\langle i, j | C \cup C' \cup \{k\} \rangle \notin \mathcal{J}$ . Since  $\mathcal{J}$  and  $G$  are faithful, there is a connecting walk  $\pi$  between  $i$  and  $j$  given  $C \cup C' \cup \{k\}$ . If  $k$  is not on  $\pi$ ,  $\pi$  is connecting given  $C \cup C'$ . Otherwise, it holds that  $k$  must be a collider on  $\pi$ . By Lemma 2 in [33],  $k$  is in ancestor of a node  $l' \in \{i, j\} \cup C \cup C'$ . By replacing the subwalk of  $\pi$  from  $l'$  to  $k$  by the directed path from  $k$  to  $l'$  if  $l' \in \{i, j\}$ , or by adding the directed path from  $k$  to  $l'$  plus the reverse from  $l'$  to  $k$  to  $\pi$  if  $l' \in C \cup C'$ , we conclude that  $\langle i, j | C \cup C' \rangle \notin \mathcal{J}$ . Hence,  $\langle i, j | C' \rangle \notin \alpha(\mathcal{J}; M, C)$ .

To prove ordered downward-stability, assume that  $\langle i, j | C' \setminus \{k\} \rangle \notin \alpha(\mathcal{J}; M, C)$ , where  $k \in C'$  is not an anterior of any node  $l \in \{i, j\} \cup C'$  in  $H$ . We have that  $\langle i, j | C \cup C' \setminus \{k\} \rangle \notin \mathcal{J}$ . Since  $\mathcal{J}$  and  $G$  are faithful, there is a connecting walk  $\pi$  between  $i$  and  $j$  given  $C \cup C' \setminus \{k\}$ . Node  $k$  cannot be on  $\pi$  since  $k$  must be a non-collider and consequently an ancestor of a node in  $\{i, j\} \cup C \cup C'$  in  $G$ , which by Lemma 3 in [33], implies that  $k \in \text{an}(l)$  in  $H$ . Hence,  $k$  is not on  $\pi$ , and consequently  $\pi$  is connecting given  $C \cup C'$ . We conclude that  $\langle i, j | C \cup C' \rangle \notin \mathcal{J}$ . Hence,  $\langle i, j | C' \rangle \notin \alpha(\mathcal{J}; M, C)$ .  $\square$

This shows that, for every ancestral graph, there is a distribution faithful to it since, for every ancestral graph, there is a DAG that, after marginalization and conditioning, results in the given ancestral graph.

**8. Summary and future work.** We provided sufficient and necessary conditions for an independence model, and consequently a probability distribution, to be faithful to a chain mixed graph. All the definitions and results concerning independence models are true for probabilistic independence models. The conditions can be divided into two main categories: 1) Those that ensure that the distribution is Markov to a graph whose skeleton is the same as the skeleton of the distribution — these properties are namely intersection, composition, and ordered upward- and downward-stabilities. If the distribution is already pairwise Markov to the graph then intersection and composition suffice. Ordered upward- and downward-stabilities direct the edges of the skeleton of the distribution so that the pairwise Markov property is satisfied. 2) Those that ensure that a Markov distribution is also faithful to the graph. These are singleton-transitivity and ordered upward- and downward-stabilities.

The type of preordering or preorderings w.r.t. which the distribution satisfies ordered upward- and downward-stabilities implies the type or types of graphs that are faithful to the distribution, which would be a subclass of antierial graphs. In order to only deal with simpler classes of graphs, it is enough to search through a more refined set of preorderings determined by the conditions that graphs of a specific class must satisfy. It is indeed true

that when the skeleton of the graph or distribution is known, the preordering is equivalent to different ways the edges of the skeleton can be directed, but in principle the preordering of the variables is unrelated to the skeleton of the distribution.

The provided conditions are clearly satisfied by the independence model of many Gaussian distributions. In general, it is important to study the characteristics of distributions that satisfy these conditions. The intersection property is well-studied and, in addition to positivity of the density that has been known to be a sufficient condition, necessary and sufficient conditions for a probabilistic independence to satisfy this property are known; see [29]. For composition and singleton-transitivity, more studies are needed; in particular, simple sufficient conditions for probability distributions in order to satisfy these properties would be beneficial.

Another important approach is to devise algorithms that test whether these conditions are satisfied given data or an independence model. Such algorithms are just some of the many algorithmic consequences the results in this paper may prompt. In particular, testing whether ordered upward- and downward-stabilities are satisfied can be combined with the intuitive algorithm that receives a probabilistic independence model induced by  $P$  as an input, and finds graphs that are faithful to  $P$  when the sufficient and necessary conditions for  $\mathcal{J}(P)$  are satisfied. We plan on providing the details of this algorithm elsewhere. This algorithm would be a *constraint-based* structural learning algorithm (as opposed to score-based, e.g., [3]) most similar to the Fast Causal Inference (FCI) algorithm [38] — it finds the skeleton of the graph and then explores directions of the edges of the graph on the skeleton. Directing the edges of the skeleton is performed based on an exploitation of the preordering of the variables (and consequently the nodes of the graph) so that w.r.t. which ordered upward- and downward-stabilities are satisfied.

When dealing with data, another issue is that the faithfulness assumption is very sensitive to hypothesis testing errors for inferring the independence statements from data; see [43]. In general, when higher order interactions exist, the existing conditional independence tests may fail; see, e.g., the Appendix in [48]. This problem is not specific to the theory in the current paper, but nevertheless, more work is needed to devise better conditional independence tests.

**Acknowledgements.** The original idea of this paper was inspired by discussions at the American Institute of Mathematics workshop “Positivity, Graphical Models, and the Modeling of Complex Multivariate Depen-



dencies” in October 2014. The author thanks the AIM and NSF for their support of the workshop and is indebted to other participants of the discussion group, especially to Shaun Fallat, Steffen Lauritzen, Nanny Wermuth, and Piotr Zwiernik for follow-up discussions. The author is also grateful to Rasmus Petersen for valuable discussions.

## References.

- [1] ALI, R. A., RICHARDSON, T. S. and SPIRITES, P. (2009). Markov Equivalence for Ancestral Graphs. *Ann. Statist.* **37** 2808–2837.
- [2] ANDERSSON, S. A., MADIGAN, D. and PERLMAN, M. D. (2001). Alternative Markov properties for chain graphs. *Scand. J. Stat.* **28** 33–85.
- [3] CHICKERING, D. M. (2003). Optimal Structure Identification with Greedy Search. *J. Mach. Learn. Res.* **3** 507–554.
- [4] CLIFFORD, P. (1990). Markov Random Fields in Statistics. In *Disorder in Physical Systems: A Volume in Honour of John M. Hammersley* (G. R. Grimmett and D. J. A. Welsh, eds.) 19–32. Oxford University Press.
- [5] COX, D. R. and WERMUTH, N. (1993). Linear dependencies represented by chain graphs (with discussion). *Stat. Sci.* **8** 204–218; 247–277.
- [6] DARROCH, J. N., LAURITZEN, S. L. and SPEED, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* **8** 522–539.
- [7] DAWID, A. P. (1979). Conditional independence in statistical theory (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **41** 1–31.
- [8] DRTON, M. and RICHARDSON, T. S. (2008). Binary models for marginal independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **41** 287–309.
- [9] FALLAT, S., LAURITZEN, S., SADEGHI, K., UHLER, C., WERMUTH, N. and ZWIERNIK, P. (2016). Total positivity in Markov structures. *Ann. Statist.* **To appear**.
- [10] FRYDENBERG, M. (1990). The chain graph Markov property. *Scand. J. Stat.* **17** 333–353.
- [11] GEIGER, D., VERMA, T. S. and PEARL, J. (1990). Identifying independence in Bayesian networks. *Networks* **20** 507–534.
- [12] GLYMOUR, C., SCHEINES, R., SPIRITES, P. and KELLY, K. (1987). *Discovering Causal Structure*. Academic Press.
- [13] KARLIN, S. and RINOTT, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Mult. Anal.* **10** 467–498.
- [14] KARLIN, S. and RINOTT, Y. (1983). M-Matrices as covariance matrices of multinormal distributions. *Linear Algebra Appl.* **52** 419 – 438.
- [15] KAUFMANN, G. (1996). On a dualization of graphical Gaussian models. *Scand. J. Stat.* **23** 105–116.
- [16] KIIVERI, H., SPEED, T. P. and CARLIN, J. B. (1984). Recursive causal models. *J. Aust. Math. Soc., Ser. A* **36** 30–52.
- [17] LAURITZEN, S. L. (1996). *Graphical Models*. Clarendon Press, Oxford, United Kingdom.
- [18] LAURITZEN, S. L. and SADEGHI, K. (2016). Unifying Markov properties for graphical models. *submitted*.
- [19] LAURITZEN, S. L. and SPIEGELHALTER, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **50** 157–224.

- [20] LAURITZEN, S. L. and WERMUTH, N. (1989). Graphical models for association between variables, some of which are qualitative and some quantitative. *Ann. Statist.* **17** 31–57.
- [21] LNĚNIČKA, R. and MATÚŠ, F. (2007). On Gaussian conditional independence structures. *Kybernetika* **43** 327–342.
- [22] MEEK, C. (1995). Strong Completeness and Faithfulness in Bayesian Networks. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence. UAI'95* 411–418. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- [23] NEAPOLITAN, R. E. (2004). *Learning Bayesian networks*. Prentice Hall, Upper Saddle River, NJ, USA.
- [24] PEÑA, J. M. (2009). Faithfulness in chain graphs: The discrete case. *Internat. J. Approx. Reason.* **50** 1306–1313.
- [25] PEÑA, J. M. (2011). Faithfulness in Chain Graphs: The Gaussian Case. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS 2011)* **15** 588–599. JMLR.org.
- [26] PEÑA, J. M. (2014). Marginal AMP Chain Graphs. *Internat. J. Approx. Reason.* **55** 1185–1206.
- [27] PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems : networks of plausible inference*. Morgan Kaufmann Publishers, San Mateo, CA, USA.
- [28] PEARL, J. (2009). *Causality: Models, Reasoning and Inference*, 2nd ed. Cambridge University Press, New York, NY, USA.
- [29] PETERS, J. (2015). On the Intersection Property of Conditional Independence and its Application to Causal Discovery. *J. Causal Inference* **3** 97–108.
- [30] RAMSEY, J., ZHANG, J. and SPIRITES, P. (2006). Adjacency-Faithfulness and Conservative Causal Inference. In *UAI*. AUAI Press.
- [31] RICHARDSON, T. (2003). Markov properties for acyclic directed mixed graphs. *Scand. J. Stat.* **30** 145–157.
- [32] RICHARDSON, T. S. and SPIRITES, P. (2002). Ancestral graph Markov models. *Ann. Statist.* **30** 962–1030.
- [33] SADEGHI, K. (2013). Stable mixed graphs. *Bernoulli* **19** 2330–2358.
- [34] SADEGHI, K. (2016). Marginalization and conditioning for LWF chain graphs. *Ann. Statist.* **44** 1792–1816.
- [35] SADEGHI, K. and LAURITZEN, S. L. (2014). Markov properties for mixed graphs. *Bernoulli*, **20** 676–696.
- [36] SCHRÖDER, B. S. W. (2003). *Ordered sets : an introduction*. Birkhäuser, Boston.
- [37] SOH, D. W. and TATIKONDA, S. C. (2014). Testing Unfaithful Gaussian Graphical Models. In *Advances in Neural Information Processing Systems 27* (Z. Ghahramani, M. Welling, C. Cortes, N. d. Lawrence and K. q. Weinberger, eds.) 2681–2689. Curran Associates, Inc.
- [38] SPIRITES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation, Prediction, and Search*, 2nd ed. MIT press.
- [39] STEEL, D. (2006). Homogeneity, selection, and the faithfulness condition. *Minds and Machines* **16** 303–317.
- [40] STUDENÝ, M. (1989). Multiinformation and the problem of characterization of conditional independence relations. *Problems of Control and Information Theory* **18** 3–16.
- [41] STUDENÝ, M. (2005). *Probabilistic Conditional Independence Structures*. Springer-Verlag, London, United Kingdom.
- [42] STUDENÝ, M. and BOUCKAERT, R. R. (1998). On chain graph models for description of conditional independence structures. *Ann. Statist.* **26** 1434–1495.
- [43] UHLER, C., RASKUTTI, G., BÜHLMANN, P. and YU, B. (2013). Geometry of the

- faithfulness assumption in causal inference. *Ann. Statist.* **41** 436–463.
- [44] VERMA, T. and PEARL, J. (1990). On the Equivalence of Causal Models. In *Proceedings of the Proceedings of the Sixth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-90)* 220–227. Elsevier Science, New York, NY.
  - [45] WERMUTH, N. (2011). Probability distributions with summary graph structure. *Bernoulli* **17** 845–879.
  - [46] WERMUTH, N. and COX, D. R. (1998). On association models defined over independence graphs. *Bernoulli* **4** 477–495.
  - [47] WERMUTH, N., MARCHETTI, G. M. and COX, D. R. (2009). Triangular systems for symmetric binary variables. *Electron. J. Stat.* **3** 932–955.
  - [48] WERMUTH, N. and SADEGHI, K. (2012). Sequences of regressions and their independences. *TEST* **21** 215–252 and 274–279.
  - [49] WOODWARD, J. (1998). Causal Independence and Faithfulness. *Multivariate Behavioral Research* **33** 129–148.
  - [50] ZHANG, J. and SPIRITES, P. (2008). Detection of Unfaithfulness and Robust Causal Inference. *Minds and Machines* **18** 239–271.

STATISTICAL LABORATORY,  
CENTRE FOR MATHEMATICAL SCIENCES,  
WILBERFORCE ROAD,  
CAMBRIDGE, CB3 0WB,  
UNITED KINGDOM.  
E-MAIL: [k.sadeghi@statslab.cam.ac.uk](mailto:k.sadeghi@statslab.cam.ac.uk)